The Annals of Applied Statistics
2021, Vol. 15, No. 1, 252–269
https://doi.org/10.1214/20-AOAS1393

MONITORING VACCINE SAFETY BY STUDYING TEMPORAL VARIATION OF ADVERSE EVENTS USING VACCINE ADVERSE EVENT REPORTING SYSTEM

BY JING HUANG[1,*], YI CAI[2], JINGCHENG DU[3,¶], RUOSHA LI[4], SUSAN S. ELLENBERG[1,†], SEAN HENNESSY[1,‡], CUI TAO[3,‖] AND YONG CHEN[1,§]

[1]*Department of Biostatistics, Epidemiology and Informatics, Perelman School of Medicine, University of Pennsylvania,* [*]*jing14@upenn.edu;* [†]*sellenbe@pennmedicine.upenn.edu;* [‡]*hennessy@pennmedicine.upenn.edu;* [§]*ychen123@upenn.edu*

[2]*AT&T Chief Data Office, AT&T Services, Inc., yc010@att.com*

[3]*School of Biomedical Informatics, University of Texas Health Science Center at Houston,* [¶]*jingcheng.Du@uth.tmc.edu;* [‖]*cui.Tao@uth.tmc.edu*

[4]*Department of Biostatistics and Data Science, University of Texas Health Science Center at Houston, Ruosha.Li@uth.tmc.edu*

The Vaccine Adverse Event Reporting System (VAERS) plays a vital role in vaccine safety surveillance. One of the main missions of VAERS is to monitor increases in reporting rate of adverse events, as such signals can indicate safety issues caused by update of vaccines or change in vaccine practices. Existing methods can rarely be used to monitor the temporal variation of reporting adverse events. In this paper we propose a composite likelihood based variance component model to study the temporal variation of reporting adverse events using VAERS data. The proposed method is devised to identify safety signals by testing the heterogeneity of reporting rates of adverse events across years. The proposed method accounts for the well-known underreporting of adverse events and the zero-inflation observations in passive surveillance reporting systems. We applied the proposed method to VAERS reports of trivalent influenza virus vaccine and identified 14 adverse events with significantly heterogeneous reporting rates over years and two of them have increasing trend of reporting rates from 1990 to 2013. Our findings provide early warning signals that can be more rigorously investigated in future studies of the vaccine.

1. Introduction. Vaccine safety is a critically important public health issue. Public confidence in vaccines depends greatly on public confidence in the government's safety surveillance. The Vaccine Adverse Event Reporting System (VAERS) is a national postmarketing passive safety surveillance program coadministered by the Centers for Disease Control and Prevention (CDC) and the Food and Drug Administration (FDA) to collect spontaneous reports of adverse events (AEs, possible side effects) that occur after the administration of any vaccine licensed in the United States. Reports are submitted by vaccine manufacturers, health care professionals, vaccine recipients and the public. From 1990 to the present, VAERS receives up to 30,000 reports annually (Centers for Disease Control and Prevention (2017c)). Like other passive surveillance systems, VAERS is subject to multiple limitations, including underreporting, recall bias, reporting errors and lack of denominator data and unbiased control groups (Ellenberg and Chen (1997)). Despite these limitations, VAERS contributes to public health in critical ways (Chen et al. (1994), Shimabukuro et al. (2015)). Because of the high number of reports and increasing national coverage, VAERS provides a unique opportunity in monitoring vaccine adverse events that might occur too rarely to be detected in prelicensure clinical trials or even postmarketing active surveillance programs (Ellenberg and

Received March 2020; revised August 2020.
Key words and phrases. Composite likelihood, heterogeneity, signal detection, underreporting, vaccine safety outcome.

Chen (1997), Thompson et al. (2018)). Effective analysis of VAERS data is vital to assuring the safety of vaccines.

One of the main missions of VAERS as well as an important aspect of vaccine safety surveillance is to monitor changes in known adverse events. Such changes can be due to many factors, for example, update of vaccines (changes in ingredients of vaccines), change in vaccine practices (addition of vaccines to existing immunization schedule, presence of newly licensed vaccines) and publicity (Chen et al. (1994), Eberth et al. (2014)). If a change (or temporal variation) of reporting AEs is identified through VAERS, investigators may conduct further studies to figure out whether the signal represents an actual risk and identify the factors that are associated with the signal. For example, a change from separate to simultaneous immunization of DTP (diphtheria, tetanus and acellular pertussis) and MMR (measles, mumps and rubella) vaccines for children at age of 15 months old could potentially change the types of AEs reported (Centers for Disease Control and Prevention (1991)). Using data from VAERS and MSAEFI (the Monitoring System for Adverse Events Following Immunization), the CDC surveillance program that preceded VAERS, investigators found that the rates of reported hospitalizations and deaths following DTP vaccine remained constant from 1985 to 1992, despite the addition of Haemophilus influenzae Type B vaccine to the routine infant immunization schedule (Chen, Haber and Mullen (1995)). These results provided some reassurance that, by adding vaccines to the recommended immunization schedules, gains in protection from disease were not offset by an increased burden of AEs caused by the additional vaccines. Another example is the change of ingredients in the manufacture of MMR in 2006, that is, human derived serum albumin was replaced by recombinant human albumin which eliminated the use of any human-derived substances (Lievano et al. (2012)). Such changes may cause the variation of reporting AEs across years. Studying temporal variation of reporting AEs is important in helping identify meaningful variation in composition of a vaccine that requires more safety analysis.

Currently, several methods have been proposed to monitor safety outcomes using data from passive surveillance programs, and these methods could potentially be applied to VAERS, given the similarity of the data structure and the research questions among the databases. For example, similar to VAERS, the FDA Adverse Event Reporting System (FAERS) is a passive postmarketing safety surveillance program that accepts AE and medication error reports for drugs and therapeutic biologic products. Methods proposed to detect safety signals using FAERS data (Greenwood and Nikulin (1996), Bate et al. (1998), DuMouchel (1999), Evans, Waller and Davis (2001), Rothman, Lanes and Sacks (2004), Huang, Zalkikar and Tiwari (2011), Huang et al. (2017), Zhao, Yi and Tiwari (2018)) can be applied to VAERS in identifying new AE signals that are associated with a given vaccine. Specifically, the proportional reporting ratios (Evans, Waller and Davis (2001)) and reporting odds ratios (Rothman, Lanes and Sacks (2004)) methods evaluate the safety of a certain vaccine/drug by calculating the proportion ratio or odds ratio of a particular AE to all the other AEs for the vaccine/drug. The Chi-square test (Pearson (1900)) can be used to test the dependence between a particular AE and the vaccination/drug (Greenwood and Nikulin (1996)). Bayesian methods were also proposed. For example, the multiitem gamma poisson shrinker method (DuMouchel (1999)) uses empirical Bayes method to test the significance of relative reporting rates of a sets of AEs by assigning a common prior to the predefined relative reporting rates. The Bayesian confidence propagation neural network method (Bate et al. (1998)) is similar to the the multiitem gamma poisson shrinker, but it uses a full Bayesian methodology.

In a recent article published at the *Journal of the American Statistical Association*, Huang, Zalkikar and Tiwari (2011) proposed a novel zero-inflated Poisson model with a likelihood ratio test to detect safety signals using FAERS data. This method has the advantage of properly accounting for the passive collection of safety reports in the surveillance program and

was shown to be powerful in identifying safety signals with well-controlled Type I errors. However, all of the aforementioned methods are effective in comparing the overall reporting rate of a particular AE to other AEs for a given vaccine/drug or comparing the overall reporting rate of a given AE for a particular vaccine/drug to other vaccines/drugs to identify safety signals. None of them aims to monitor the temporal variation of reporting AEs which is a critical question in monitoring vaccine safety.

To study the temporal variation of reporting AEs using VAERS data, several aspects need to be considered. First, the total number of vaccine recipients and the total number of vaccine recipients who are willing to report in VAERS cannot be determined from VAERS data. The changes in number of reported AEs can be caused by the fluctuation of the unknown population over time and among many other reasons (Chen et al. (1994), Haber et al. (2004)). For example, number of reports for a new vaccine are always large when it is first put on the market and may diminish over time. This pattern is also known as the "Weber effect" (Weber (1984)). Public events and social media reports that raise publicity can also lead to a burst of new reports which is also known as "stimulated reporting" (Eberth et al. (2014)). In addition, if the indication for a vaccine is expanded to a broader population group (e.g., a vaccine approved in toddlers later approved in infants), the number of reported events tends to increase (Ellenberg and Chen (1997)). Second, VAERS data can rarely provide definitive evidence of causal relationships between a vaccine and a particular AE due to underreporting, recall bias, reporting errors and lack of denominator data and unbiased control groups (Singleton et al. (1999), Zhou et al. (2003)). However, this type of national reporting system can rapidly document the possibility of unexpected AEs, generating early warning signals that can then be more rigorously investigated in focused studies, including Vaccine Safety Datalink (Chen et al. (1997), McNeil et al. (2014)). In a sense, VAERS is the front line of vaccine safety surveillance, so methods developed for VAERS should give sensitivity precedence over specificity.

In this paper, built upon the work of Huang, Zalkikar and Tiwari (2011), we propose a composite likelihood variance component model to study the temporal variation of reporting AEs with the goal of detecting safety signals using VAERS data. To the best of our knowledge, the proposed method is the first to study the temporal variation of reporting AEs using VAERS data. Our method accounts for the unique features of VAERS data and has several advantages. First, we model the likelihood of reporting a given AE conditional on the total number of reports observed in each year. Such a conditional likelihood method alleviates the impact of fluctuation of the vaccinated population in comparing reported AEs across years. Second, we use a parsimonious composite likelihood method to reduce the complexity of the model which was devised to be sensitive to signals in fluctuation in time. Such a composite likelihood method also has better computational performance and model robustness; for examples, see Chen et al. (2014, 2015). For further discussion of the composite likelihood method and its applications, we refer to a review paper by Varin, Reid and Firth (2011) and the references therein. Moreover, the proposed method accounts for the zero-inflated feature of VAERS data.

The rest of this paper is organized as follows. In Section 2 we introduce a motivating example of using the VAERS dataset for monitoring the safety signals for trivalent influenza virus vaccine (FLU3). In Section 3 we describe the proposed composite likelihood ratio test. The asymptotic distribution of the proposed likelihood ratio test is derived. In Section 4 we conduct simulation studies to compare the performance of the proposed test with existing tests in terms of Type I errors and power. In Section 5 we apply the proposed method to the VAERS FLU3 dataset to detect safety signals. Finally, we provide a discussion on the strengths and limitations of the proposed method in Section 6.

2. Monitoring FLU3 safety using VAERS data. Influenza is a highly contagious viral respiratory infection that affects 5–20% of the U.S. population each year (Centers for Disease Control and Prevention (2016a)). More than 140,000 individuals are hospitalized and 12,000–56,000 die from influenza-related complications in the United States for the 2015–2016 influenza season (Centers for Disease Control and Prevention (2017a)). The best way to prevent influenza is to receive an annual influenza vaccination (Centers for Disease Control and Prevention (2017b), Grohskopf et al. (2015)). The CDC recommend that everyone at six months of age and older to receive influenza vaccination every year (Centers for Disease Control and Prevention (2017b)).

Among all the VAERS reports, FLU3 is the most common vaccine type reported, accounting for nearly 12% of the entire VAERS database. Thus, VAERS is potentially a very important data source for monitoring FLU3 safety. FLU3 is a synthetic influenza vaccine consisting of three inactivated influenza viruses, including an influenza A H1N1 virus, an influenza A H3N2 virus and a B virus (Centers for Disease Control and Prevention (2016b)). Influenza vaccine safety using VAERS data has been studied extensively (Haber et al. (2004), Vellozzi et al. (2009)), but none of this work has taken into account an important feature of influenza vaccines: most influenza vaccines are formulated annually based on influenza strains projected to be prevalent in the upcoming flu season which can lead to changes in types and numbers of reported AEs. As we will demonstrate later, the VAERS reports for FLU3 are of high prevalence and with substantial variations over the years.

Upon searching the AE database for FLU3 in the United States from 1990 to 2013, we extracted 6813 reports with at least one of the following serious reactions: death, life-threatening illness, hospitalization, prolonged hospitalization or permanent disability. All of these reports were manually coded using Medical Dictionary for Regulatory Activities (MedDRA) vocabulary by domain experts (MedDRA (2017)). The VAERS reports of FLU3 from 1990 to 2013 include 3784 unique Preferred Terms (PTs). Each PT is "a distinct descriptor (single medical concept) for a symptom, sign, disease diagnosis, therapeutic indication, investigation, surgical or medical procedure and medical social or family history characteristic" (MedDRA (2017)). Based on the hierarchical structure of MedDRA, we mapped the PTs to 26 System Organ Classes (SOCs), which is the highest level of the hierarchy in MedDRA that comprises grouping by etiology, manifestation site and medical purpose, using a method developed by Du et al. (2016).

A primary analysis of the data showed, in Figure 1, that the total number of occurrence of AEs reported to VAERS for FLU3 ranged from 162 to 11,760 each year. The overall trend

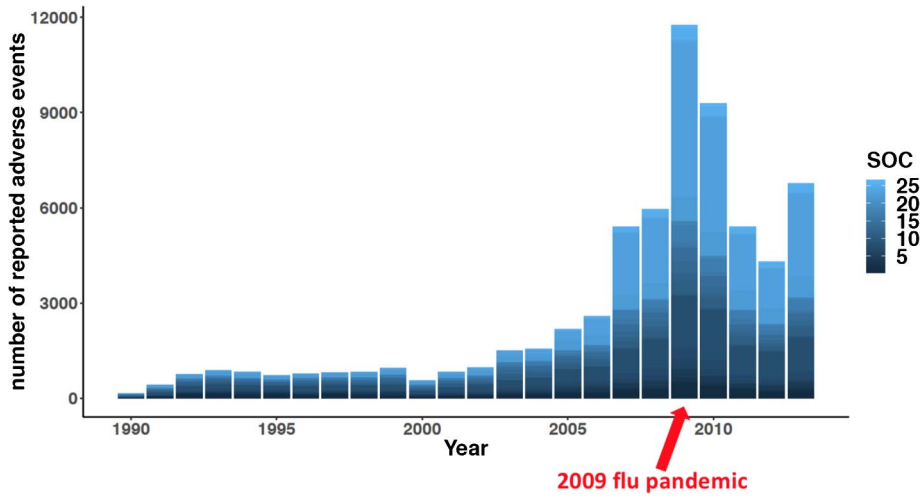

FIG. 1. *Total number of AEs reported to VAERS for FLU3 from* 1990 *to* 2013 *by the category of SOCs.*

is increasing from 1990 to 2013, especially after 2005, potentially due to the increasing national publicity of VAERS and public awareness of reporting AEs. In 2009 and 2010, VAERS received a peak of AE reports for FLU3, partially due to the emergence of the 2009 H1N1 influenza virus in the spring of 2009 and an increasing number of vaccinated people during the 2009–2010 influenza season (Goldman (2013)). Given this variability, monitoring of AEs for FLU3 using VAERS data presents a particular challenge that the number of AE reports varies from year to year due to the fluctuation in total number of vaccinated population, the length and severity of influenza seasons and publicity factors. It is hence of substantial interest to develop a model that can effectively use VAERS data to monitor any temporal variation of AE reporting rates for vaccine safety.

3. Proposed method.

3.1. *Notations, proposed models and model assumptions.* To study the temporal variation of reporting AEs, VAERS reports from multiple years for a particular vaccine, for example, FLU3, can be structured as an $I \times J$ table with a total of I reporting years being the rows and J types of AE being the columns. The count for the (i, j)th cell, n_{ij}, is the number of events reported for the jth AE in the ith year for the vaccine. The total number of reported events in the ith year is the marginal total of the ith row denoted as $n_{i\cdot}$, the total number of reported events for the jth AE over all the I years is the marginal total of the jth column, denoted as $n_{\cdot j}$, and the total number of all reported AEs for the vaccine over all the years is the grand total, denoted as $n_{\cdot\cdot}$.

As mentioned earlier, one limitation of VAERS data is that the total number of vaccine recipients, the total number of vaccine recipients who are willing to report to VAERS and the total number of events that actually occurred are unknown. Thus, a simple comparison of numbers of events reported across years is confounded by the fluctuation of unknown population size over time. To circumvent this problem, we model the number of events reported for the jth AE in the ith year, n_{ij}, conditioning on the total number of events reported in the ith year, $n_{i\cdot}$, which can be assumed to follow a Binomial distribution, even though the total underlying population in each year is unknown, that is, $n_{ij} \sim \text{binomial}(n_{i\cdot}, p_{ij})$, where p_{ij} is the rate of reporting the jth AE to all the AEs in the ith year.

Another feature of the VAERS data, as well as data from most surveillance reporting systems, is that they are usually sparse, that is, $n_{i\cdot}$ is large and n_{ij} is small, and contain a large number of zero-count cells, that is, $n_{ij} = 0$ for many (i, j) pairs. In the VAERS database with more than 4000 types of AEs, the percentage of vaccines that have more than 90% of observed zero-count cells is as large as 88%. The percentage of zero values ranges from 57% to 99%. To account for these features, we use a Poisson approximation to the binomial distribution with computational advantage and propose a zero-inflated Poisson model as follows:

$$n_{ij} = \begin{cases} 0 & \text{with probability } w_j, \\ \sim \text{Poisson}(n_{i\cdot} p_{ij}) & \text{with probability } 1 - w_j, \end{cases} \tag{3.1}$$

where w_j is the probability of excess zeros for the jth AE. The model assumes that the observed zero counts can be generated from two different mechanisms, that is, one follows a Bernoulli distribution with probability w_j and the other follows a Poisson distribution with mean $n_{i\cdot} p_{ij}$. The two parameters, w_j and p_{ij}, are estimated simultaneously. Given w_j, $n_{i\cdot}$ and p_{ij}, the observation n_{ij} follows a zero-inflated Poisson distribution.

In order to study the temporal variation of reporting rate of the jth AE over years, we further assume a variance component model to allow the heterogeneity of reporting rates, such that

$$\begin{aligned} \text{logit}(p_{ij}) &\sim \text{N}(\beta_{0j}, \tau_j^2), \quad i = 1, \ldots, I, \\ \beta_{0j} &= \text{logit}(p_{0j}), \end{aligned} \tag{3.2}$$

where β_{0j} is the average reporting rate of the jth AE across all years in the logit scale and τ_j^2 quantifies the heterogeneity of reporting rates of the jth AE across years.

3.2. *A composite likelihood-based variance component test for studying temporal variation of reporting AEs.* Given the proposed model (3.1) and (3.2), our primary interest is to identify AEs with increased reporting rates at certain years for a particular vaccine. Statistically, this can be done by testing the heterogeneity of reporting rates across years, which is equivalent to testing the null hypothesis, $H_0 : \tau_j^2 = 0$, for the jth AE.

To construct a likelihood function for the proposed model (3.1) and (3.2), a major challenge is that the numbers of events reported for the jth AE for the same vaccine across years are correlated and the correlation structure is usually complicated and unknown, as a group of people can get vaccinated in consecutive years and may have similar AEs postvaccination. Given such complex correlations in AEs across years, instead of formulating a joint distribution on AEs over years, we propose to construct a composite likelihood by multiplying individual conditional likelihood together across years (Lindsay (1988)). Such a modeling strategy is parsimonious by avoiding modeling the complex correlations among AEs across years. Specifically, under the null hypothesis, $H_0 : \tau_j^2 = 0$, a composite likelihood function using data related to the jth AE can be written as

$$
\begin{aligned}
&L_0(w_j, p_{0j}; n_{ij}, n_{i.}) \\
&\quad = \prod_{i=1}^{I} \{w_j + (1 - w_j)\exp(-n_{i.}p_{0j})\}^{u_i} \\
&\qquad \times \{(1 - w_j)\exp(-n_{i.}p_{0j})(n_{i.}p_{0j})^{n_{ij}}/n_{ij}!\}^{1-u_i},
\end{aligned} \tag{3.3}
$$

where $u_i = I(n_{ij} = 0)$. Notably, the marginal densities of $n_{1j}, \ldots, n_{Ij}$ are multiplied together without accounting for the correlation among them to construct a composite likelihood function (Cox and Reid (2004), Lindsay (1988), Varin, Reid and Firth (2011)). Under the alternative hypothesis, $H_a : \tau_j^2 > 0$, a composite likelihood function can be constructed as

$$
\begin{aligned}
&L_a(w_j, \beta_{0j}, \tau_j^2; n_{ij}, n_{i.}) \\
&\quad = \prod_{i=1}^{I} \int_0^1 p(n_{ij}|n_{i.}; p_{ij}, w_j) p\big(p_{ij}|p_{0j}, \tau_j^2\big)\, dp_{ij} \\
&\quad = \prod_{i=1}^{I} \int_0^1 \{w_j + (1 - w_j)\exp(-n_{i.}p_{ij})\}^{u_i} \\
&\qquad \times \{(1 - w_j)\exp(-n_{i.}p_{ij})(n_{i.}p_{ij})^{n_{ij}}/n_{ij}!\}^{1-u_i} \\
&\qquad \times \{p_{ij}(1 - p_{ij})\tau_j\sqrt{2\pi}\}^{-1} \exp\left[-\frac{\{\mathrm{logit}(p_{ij}) - \beta_{0j}\}^2}{2\tau_j^2}\right] dp_{ij}.
\end{aligned} \tag{3.4}
$$

Denoting the maximum composite likelihood estimates under the null and the alternative hypothesis as $(\hat{w}_j, \hat{p}_{0j})$ and $(\tilde{w}_j, \tilde{\beta}_{0j}, \tilde{\tau}_j^2)$, respectively, a composite likelihood ratio test for variance component, that is, testing the heterogeneity of the reporting rates of the jth AE for a given vaccine across years, can be constructed as

$$
LR_j = \frac{L_a(\tilde{w}_j, \tilde{\beta}_{0j}, \tilde{\tau}_j^2; n_{ij}, n_{i.})}{L_0(\hat{w}_j, \hat{p}_{0j}; n_{ij}, n_{i.})}. \tag{3.5}
$$

Under the null hypothesis, the parameter of interest τ_j^2 lies on the boundary of the parameter space, and the limiting distribution of LR_j is a 50:50 mixture of χ_0^2, that is, a point mass at 0 and weighted χ_1^2 distributions (Chen and Liang (2010), Chen et al. (2017), Huang et al. (2020)),

$$LR_j \to \frac{1}{2}\chi_0^2 + \frac{1}{2}\frac{e^*}{e}\chi_1^2, \tag{3.6}$$

where e^* and e are the elements in the inverse of Godambe information matrix Godambe (1960), $E\{-\frac{\partial^2 \log L_0}{\partial(\tau_j^2)^2}\}\{\text{var}\,\frac{\partial \log L_0}{\partial \tau_j^2}\}^{-1} E\{-\frac{\partial^2 \log L_0}{\partial(\tau_j^2)^2}\}$ and sensitivity matrix, $E\{-\frac{\partial^2 \log L_0}{\partial(\tau_j^2)^2}\}$ which can be calculated empirically. By comparing the test statistics with the critical values of the mixture χ^2 distribution, we can identify the AEs with significantly heterogeneous reporting rates across years at prespecified significance level.

3.3. *Profiling the temporal variation of reporting AEs with heterogeneous reporting rates using empirical Bayes estimator.* If the null hypothesis is rejected for the jth AE by the proposed test described in Section 3.2, we conclude that the reporting rates of the jth AE for the vaccine are statistically different across years. Such a temporal variation can indicate safety issues. To further investigate the variation and identify the factors that may be associated with it, we propose an empirical Bayes estimator to identify the years with significantly higher reporting rates of the jth AE. Different from the naive estimation of reporting rate using $\hat{p}_{ij} = n_{ij}/n_{i.}$, $i = 1, \ldots, I$, the empirical Bayes estimator borrows information across years to achieve higher accuracy with lower mean squared error (Clayton and Kaldor (1987)).

Specifically, by equation (2), for the jth AE the reporting rates share a common distribution as

$$\text{logit}(p_{ij}) \sim \text{N}(\beta_{0j}, \tau_j^2), \quad i = 1, \ldots, I,$$

where β_{0j}, τ_j^2 are the hyperparameters that can be empirically calculated from the data. The posterior distribution of reporting rate in the ith year, p_{ij}, can be written as

$$\begin{aligned} &f(p_{ij}|n_{ij}, n_{i.}) \\ &\quad \propto f(n_{ij}|p_{ij}, n_{i.}, w_j) f(p_{ij}|\beta_{0j}, \tau_j^2) \\ &\quad = \{w_j + (1-w_j)\exp(-n_{i.}p_{ij})\}^{u_i} \{(1-w_j)\exp(-n_{i.}p_{ij})(-n_{i.}p_{ij})^{n_{ij}}/n_{ij}!\}^{1-u_i} \\ &\qquad \times \{p_{ij}(1-p_{ij})\tau_j\sqrt{2\pi}\}^{-1} \exp\left[-\frac{\{\text{logit}(p_{ij}) - \beta_{0j}\}^2}{2\tau_j^2}\right], \end{aligned} \tag{3.7}$$

and the posterior mean is a shrinkage estimator of p_{ij} which borrows information across all the years. To numerically estimate the posterior mean, we use a Gibbs sampler to draw samples from equation (3.7). In each iteration the parameters w_j, β_{0j} and τ_j^2 are replaced by the maximum composite likelihood estimates of $L_a(w_j, \beta_{0j}, \tau_j^2; n_{ij}, n_{i.})$, $(\tilde{w}_j, \tilde{\beta}_{0j}, \tilde{\tau}_j^2)$ which makes the posterior mean an empirical Bayes estimator of p_{ij}.

To identify the years that have disproportionately higher reporting rates of the jth AE, we propose to rank the posterior means of $p_{ij}s$, $i = 1, \ldots, I$. We can also visualize the temporal variation of reports of the jth AE after the vaccination by plotting the posterior mean of $p_{ij}s$ vs. years.

4. Simulation study.

4.1. *Testing of heterogeneity of reporting rates of AE over years.* We conducted simulation studies to evaluate the type I error and the power of the proposed test. From the VAERS

TABLE 1

Empirical rejection rates (%) in 1000 simulations of the proposed method to test the heterogeneity of reporting rates of AE over $I = 25, 50$ years, with the heterogeneity of the reporting rates τ_j^2 varying from 0 to 0.1, and the probability of observing zero event w_j increasing from 0 to 0.8

	$I = 25$					$I = 50$				
	τ_j^2					τ_j^2				
w_j	0	0.025	0.05	0.075	0.1	0	0.025	0.05	0.075	0.1
0	3.8	26.2	60.3	82.9	89.0	4.9	51.9	86.0	96.6	99.5
0.2	3.4	25.5	52.6	74.5	81.9	3.7	44.8	77.8	95.2	97.6
0.4	3.8	21.7	44.9	60.8	69.5	3.9	35.3	67.2	87.9	93.4
0.6	1.7	14.7	30.6	45.6	50.6	2.1	27.1	51.9	74.9	82.7
0.8	1.6	6.1	15.3	24.6	30.9	2.9	12.2	28.3	46.4	57.1

data for FLU3 (details of the data are discussed in Sections 2 and 5), we simulate $I \times J$ tables with the row variable as calendar years and the column variable as types of AE after vaccination. The marginal counts for each year/row are simulated from a uniform distribution $U(1000, 5000)$ which is similar to the real data. The cell counts, that is, numbers of reported events for each type of AE in each year, are generated using the zero-inflated Poisson model in equation (3.1) given the marginal counts for each year. To evaluate the type I error, the jth columns of the data are simulated from the null hypothesis with the reporting rates assumed to be homogeneous across years, that is, $p_{ij} = p_{0j}$, $i = 1, \ldots, I$, thus the value of τ_j^2 equals to zero in equation (3.2). To evaluate the power of the proposed test, the jth column of the data are simulated under the alternative hypothesis with reporting rates simulated from equation (3.2) with $\tau_j^2 > 0$. We set the average reporting rate $p_{0j} = 0.005$ and investigate various scenarios with different probabilities of zero events, w_j, different magnitudes of heterogeneity of reporting rates, τ_j^2, and different number of years, I, to examine the performance of the proposed test in different settings. The simulation is replicated for 1000 types of AEs, that is, $J = 1000$. The VAERS reports currently date from 1990 to present. In order to evaluate the performance of the proposed test when the number of years increases, we investigate the scenarios of $I = 25$ and 50.

Table 1 summarizes the simulation results. We observe that the proposed test can control type I error very well when the number of years gets larger, for example, $I = 50$, and the probability of zero events gets lower, for example, $w_j < 0.4$. Otherwise, the proposed test is relatively conservative, for example, when $I = 25$, or $w_j = 0.6$. The proposed test has reasonable power to identify heterogeneity of reporting rates across years, even when the variance of the reporting rate is as low as 0.025. The power of the proposed test increases when number of years increases and when the probability of zero events decreases.

In summary, the simulation results suggest that the proposed test controls the type I error well with data observed from a moderate number of years and has the power to identify small heterogeneity of reporting rates.

4.2. *Empirical Bayes estimator.* We also compare the accuracy of estimating the rank of the reporting rate across years by using the naive method and empirical Bayes method. The naive method directly estimate the reporting rate of each year by $p_{ij} = n_{ij}/n_{i.}$. The empirical Bayes method estimates the parameters of the prior distribution of p_{ij} using the observations first and then calculate the posterior mean of p_{ij} given the data. The accuracy of the rank

TABLE 2

Comparison of the accuracy of rank estimation between the naive method and the proposed empirical Bayes method based statistics R_1, R_3 and ρ in 1000 simulations with $p_{0j} = 0.005$ and the number of years $I = 25$. .eb and .naive represent the proposed empirical Bayes estimator and the naive estimator, respectively

w_j	τ_j^2	R_1.eb	R_1.naive	R_3.eb	R_3.naive	ρ.eb	ρ.naive
0	0.025	0.558	0.554	0.102	0.09	0.831	0.828
	0.05	0.68	0.676	0.162	0.158	0.895	0.894
	0.075	0.732	0.73	0.218	0.222	0.923	0.922
	0.1	0.782	0.776	0.278	0.264	0.939	0.938
0.2	0.025	0.53	0.532	0.09	0.066	0.828	0.826
	0.05	0.662	0.666	0.15	0.138	0.893	0.892
	0.075	0.734	0.728	0.224	0.218	0.920	0.919
	0.1	0.768	0.774	0.288	0.266	0.934	0.934
0.4	0.025	0.548	0.56	0.106	0.11	0.821	0.823
	0.05	0.662	0.68	0.186	0.174	0.888	0.887
	0.075	0.738	0.734	0.25	0.238	0.915	0.914
	0.1	0.784	0.778	0.314	0.3	0.930	0.929
0.6	0.025	0.596	0.598	0.132	0.15	0.800	0.809
	0.05	0.686	0.682	0.242	0.25	0.872	0.875
	0.075	0.718	0.72	0.31	0.308	0.901	0.902
	0.1	0.752	0.746	0.364	0.354	0.918	0.918

estimation is characterize by several statistics:

$$R_1 = \frac{\text{Number of times correctly identify the year with largest reporting rate}}{\text{Number of replications}},$$

$$R_3 = \frac{\text{Number of times correctly identify the year with first three largest reporting rate}}{\text{Number of replications}},$$

$\rho =$ Spearman correlation between the true rank and estimated rank.

As shown in Table 2, the rank estimation by using the empirical Bayes method has a slightly better accuracy than the naive method.

In Appendix B we also provided results from additional simulation scenarios with $I = 100$ and rejection rates evaluated at the nominal levels of 0.01 and 0.0025. Our simulations and the data analyses are conducted using R version 3.5.2 (see Supplementary Material (Huang et al. (2021))), and the codes will be made available at https://github.com/Penncil/.

5. Application to VAERS FLU3 data. We applied the proposed test to the VAERS FLU3 data described in Section 2. We first conducted the analysis at the SOC level of adverse events. The total number of SOCs is 26, and the number of years studied is 24. A list of the 26 SOCs is shown in Table 3. After the mapping the data is a 24×26 table with 74 zero-count cells (11.9%).

We tested the heterogeneity of reporting rates over the 24 years for each of the 26 AEs (in terms of SOCs) using the proposed composite likelihood ratio test. Figure 2 summarizes the p-values, together with the total number of events reported during the study period for each SOC. We observed a separation of p-values of the 26 SOCs, indicated by the red solid line with a significance level of 10^{-6} in Figure 2. There were eight SOCs that had significantly heterogenous reporting rates over years ($p \leq 10^{-6}$) than the other 18 SOCs ($p > 10^{-6}$), and 10 of the 26 SOCs had significantly heterogenous reporting rates at significance level 0.05 after Bonferroni correction, indicated by the blue dash line in Figure 2.

TABLE 3
The MedDRA Terminology SOC List

SOC1	Infections and infestations
SOC2	Neoplasms benign, malignant and unspecified (including cysts and polyps)
SOC3	Blood and lymphatic system disorders
SOC4	Immune system disorders
SOC5	Endocrine disorders
SOC6	Metabolism and nutrition disorders
SOC7	Psychiatric disorders
SOC8	Nervous system disorders
SOC9	Eye disorders
SOC10	Ear and labyrinth disorders
SOC11	Cardiac disorders
SOC12	Vascular disorders
SOC13	Respiratory, thoracic and mediastinal disorders
SOC14	Gastrointestinal disorders
SOC15	Hepatobiliary disorders
SOC16	Skin and subcutaneous tissue disorders
SOC17	Musculoskeletal and connective tissue disorders
SOC18	Renal and urinary disorders
SOC19	Pregnancy, puerperium and perinatal conditions
SOC20	Reproductive system and breast disorders
SOC21	Congenital, familial and genetic disorders
SOC22	General disorders and administration site conditions
SOC23	Investigations
SOC24	Injury, poisoning and procedural complications
SOC25	Surgical and medical procedures
SOC26	Social circumstances

To further identify the SOCs that had significantly varying reporting rates, we used the proposed empirical Bayes method to estimate the reporting rate in each year for the 10 SOCs with significant p-values. We observed the trajectories of the reporting rates over time could be summarized to three categories: fluctuation without an obvious trend, decreasing trend and

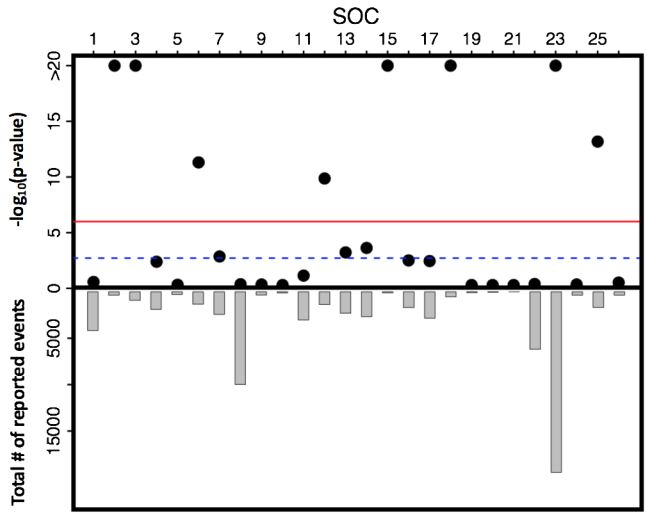

FIG. 2. *P-values obtained from the proposed test in testing the heterogeneity of reporting rates of AEs during* 1990 *to* 2013 *and total number of reported AEs for FLU*3 *from* 1990 *to* 2013. *The blue dash line indicates a significance level of* 0.05 *after Bonferroni correction, and red solid line indicates a significance level of* 10^{-6}.

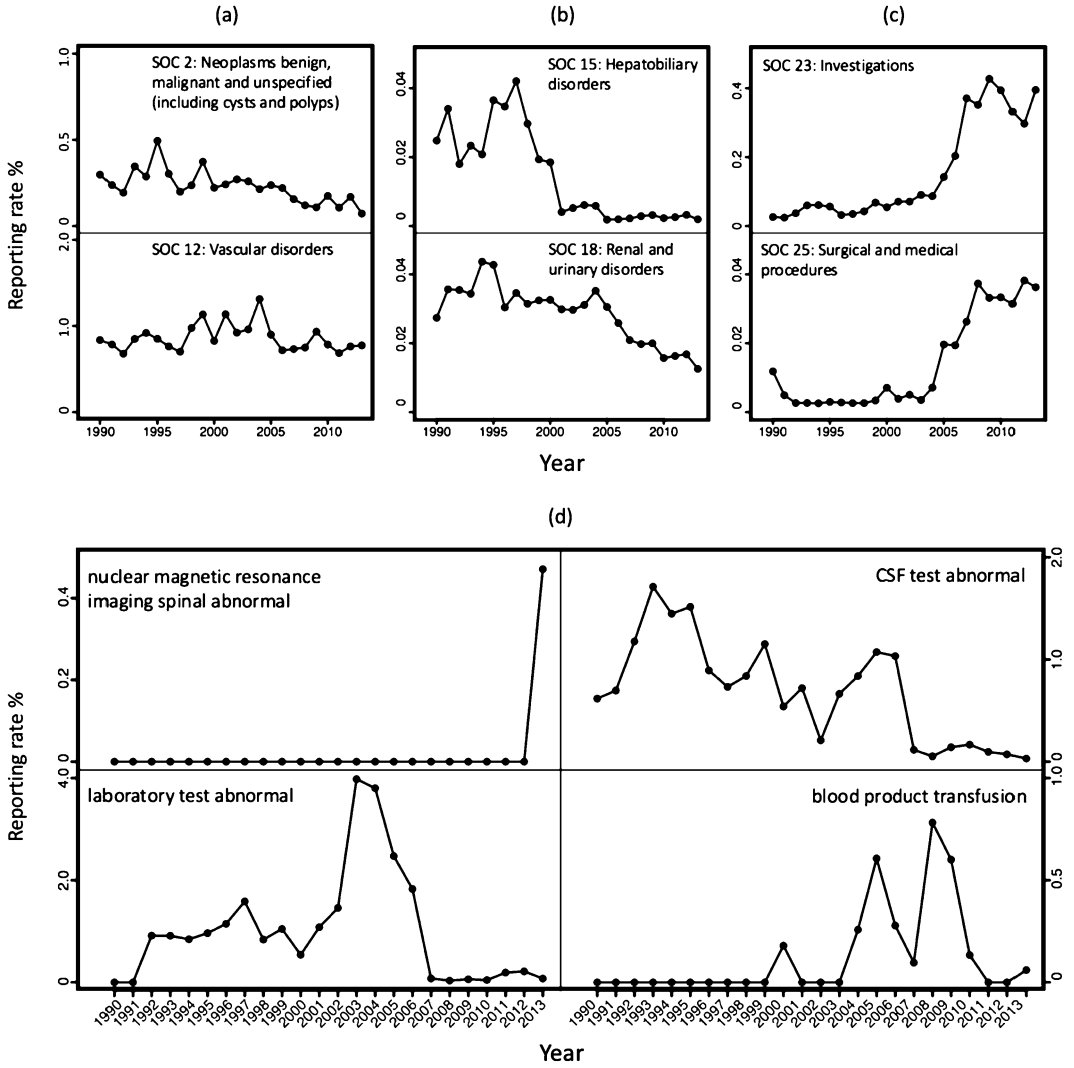

FIG. 3. *Examples of AEs that had significantly varying reporting rates from* 1990 *to* 2013 *at the SOC and PT levels*. (*a*)–(*c*): *Examples of AEs at the SOC level with reporting rate that had* (*a*) *yearly fluctuations without an obvious trend*, (*b*) *decreasing trends and* (*c*) *increasing trends*. (*d*): *Examples of AEs at the PT level within SOCs* 23 *and* 25 *that had significantly heterogeneous reporting rates*.

increasing trend. For each category we showed two examples in the column panels of Figure 3(a)–(c). Specifically, to gain further insights into FLU3 vaccine safety, we zoomed into the PT level of AEs and applied the proposed test to all the PTs that are included in SOC 23 and 25 which showed increasing reporting rates. We found four PTs had significantly heterogeneous reporting rates from 1990 to 2013 in SOC 23 and 25: nuclear magnetic resonance imaging spinal abnormal, CSF test abnormal, laboratory test abnormal and blood product transfusion. We visualized the observed reporting rates of these PTs in each year (number of VAERS reports, including the PT after taking FLU3 vaccine/ total number of VAERS reports related to FLU3 vaccine) in Figure 3(d).

We found the nuclear magnetic resonance imaging spinal abnormal was not reported until 2013. In 2013, there were 31 reports (0.5% of total number of VAERS reports related to FLUE3). By investigating the MedDRA in these consecutive years, we found such increase was due to the update of PT in MedDRA. The PT, the nuclear magnetic resonance imaging spinal abnormal, was not included in MedDRA until version 16.0 was released in March

2013. Thus, this signal could be a false positive caused by the update of MedDRA vocabulary. The second PT identified is CSF test (cerebrospinal fluid test), which is commonly used for diagnosis of conditions that affect the brain and spinal cord, including diagnosis of autoimmune disorders, such as Guillain–Barré syndrome (GBS). In Figure 3(d) we detected two peaks of reporting CSF test after FLU3 vaccination, 1993–1995 and 2003–2006, suggesting potential increasing risk of GBS related to FLU3 vaccine in these years. Both signals were confirmed, investigated and reported in literatures (Lasky et al. (1998), Iqbal et al. (2015)). Lasky et al. (1998) conducted investigation using medical data from hospital-discharge summaries in four states to study the potential increase of influenza vaccine associated GBS during 1992–1994, and similar investigation was conducted by Iqbal et al. (2015) in 2003–2006.

Our method also identified heterogeneous reporting rates of laboratory test abnormal and blood product transfusion from 1990 to 2013. Specifically, the VAERS received increased numbers of reports with laboratory test abnormal in 2002–2006 and with blood product transfusion in 2004–2006 and 2008–2009. Through discussions with a group of investigators at CDC who leads VAERS study, we found that the laboratory test abnormal and blood product transfusion are generic descriptions which can contain a combination of AEs. Since the use of the general terms can be affected by other more specific terms being available, it would potentially be beneficial to explore the wider range of more specific lab test or transfusion term changes during the period of signals. However, as VAERS reports do not contain the lower level term (LLT) of AEs, we were not able to identify the specific AE types that were driving the spike in reporting rates.

6. Discussion. In this paper we proposed a variance component test to study the temporal variation of reporting AEs over years for vaccine safety using VAERS data. To the best of our knowledge, this method is the first effort to fill the methodological gap in rigorously monitoring vaccine safety via testing the temporal variation of reporting AEs in VAERS. Our simulation studies demonstrated that the proposed method can control type I error and is sensitive in detecting temporal variation of reporting rate of AE over years in various scenarios. We applied the proposed method to the FLU3 data from VAERS and found 10 SOCs had significantly heterogenous reporting rates from 1990 to 2013, and two of them were increasing. Within the two SOCs we detected four PTs with increasing reporting rates from 1990 to 2013. In this study we illustrated the proposed method using the example of testing for annual variation of AE reports of FLU3 vaccine. In practice, the safety surveillance may need to occur more frequent than annual analysis. The proposed method can be adapted to test for monthly or quarterly heterogeneity as needed.

The proposed method has several strengths. First, our method attempted to address several key limitations of VAERS data, including the unavailability of the whole vaccinated population, underreporting of AEs and zero-inflated observations. Specifically, the changes in number of reported AEs can be caused by the fluctuation of total number of vaccinated people who are willing to report in VAERS over time. A simple comparison of the total counts of AEs reported in VAERS over years can be confounded by such an unknown fluctuation which may be due to many other reasons, such as public awareness. The proposed method alleviates the impact of fluctuation of unknown population in studying variation of reporting AEs across years using the conditional likelihood. Second, the proposed method uses a composite likelihood approach to achieve parsimony of the statistical model. Specifically, the numbers of AEs reported for a particular vaccine across years are correlated, but the correlation structure among the consecutive AE reports is complicated, as the vaccinated population vary over year and are largely unknown. The proposed composite likelihood method is appealing in this scenario, as it circumvents the challenge of modeling such a complex correlation structure,

which is commonly not of primary interest, yet provides valid statistical inference. Third, the proposed method is powerful in detecting positive signals by using a variance component test to test the heterogeneity of reporting rates over years. Moreover, the proposed method accounts for the zero-inflated feature of VAERS data and quantify the temporal variation of the reporting rate based on a ranking procedure of the estimated reporting rates over years by using empirical Bayes method and the shrinkage estimator.

Our approach also has a few limitations that deserve further investigation. The proposed method can be used by investigations as a filtering method to detect safety signals, but it cannot be used as a significance test to quantify the strength of the evidence of heterogeneity. The magnitude of the p-value obtained from the test should not be overinterpreted. It suggests statistical significance which is not an indicator of the importance of the evidence or a true causal effect. Larger sample size (number of years) can achieve higher statistical significance, but the clinical meaning of the results should be discussed with domain experts. Second, the proposed method may have low power to detect safety signals for vaccines that are not frequently updated or safety signals that are caused by reasons other than the update of vaccines or changes of practices.

Our method is a tool to detect temporal variation of reporting adverse events for existing vaccines. For new vaccines that are approved for a short period of time, there may be insufficient years of data to detect such temporal variation. In this situation, methods that comparing the overall reporting rate of a particular AE to other AEs for a given vaccine or comparing the overall reporting rate of a given AE for a particular vaccine to other vaccines, for example, the proportional reporting ratios and reporting odds ratios, are more appropriate. In event of unexpected pandemics, like the outbreaks of COVID-19 in 2020, the total number of AE reports in VAERS may be substantially different from other years. The advantage of our method is that we model the likelihood of reporting a given AE conditional on the total number of reports observed in each year. Such a conditional likelihood method alleviates the impact of fluctuation of the vaccinated population in comparing reported AEs across years.

In our investigation the proposed method also detected false positive signals caused by update of MedDRA vocabulary. MedDRA updates two times a year; each includes a combination of changes to existing terms and new terms added to MedDRA (MedDRA (2017)). Signals of increased reports can be due to the addition of new terms or merging of multiple terms into one. This makes the study of temporal variation challenging. However, as noted by CDC, the study of VAERS is deemed as safety signal detection and hypothesis generation rather than confirmatory data analysis. We believe the proposed method is a useful procedure, which is sensitive to true signals and reduces false positive detection associated with increased number of total VAERS reports, that can be attributable to factors like publicity and increased coverage of vaccination population. In this study we scanned safety signal at the SOC level and then zoomed into the PT level. Alternatively, we may scan all signals at the PT level. However, such strategy is more affected by the changing of PT definitions in MedDRA and suffers more from the multiple tests. Empirical comparisons of these strategies are important and instructive to practical investigators. This topic will be investigated in the future.

As a passive database for epidemiological studies, VAERS has several limitations. One major problem is the unavailability of a control group because adverse events in unvaccinated people are not reported to VAERS. Thus, it is difficult to assess whether the number of reported events is different from the number that would have been observed in the absence of vaccination. One possible way to address this issue may be combining the VAERS data with data from the Vaccine Safety Datalink project (Centers for Disease Control and Prevention (2017d)), where electronic health data and the health status of vaccine recipients are

available. The quality of VAERS data is also less than optimal. Because reports are submitted by a wide variety of individuals, few of whom are experienced in completing data forms for medical studies, many reports omit important data and contain obvious errors. Given that VAERS receives over 30,000 reports annually, no attempt is made to assure the accuracy and completeness of the database, although checks and follow-up are performed for a few key data items, such as the type of vaccine administered and the severity of the event. Finally, the simultaneous administration of multiple vaccines, following currently recommended vaccine schedules, further complicates the assessment of AEs because it is challenging to determine which of the vaccines (if any) was most likely to cause the outcome.

Due to these limitations, VAERS data are typically unable to provide definitive evidence of causal associations between vaccines and particular reported outcomes. Nevertheless, VAERS contributes to public health in critical ways. For example, reports in VAERS can rapidly document possible effects, cover a larger population, longer follow-up time and contain more types of AEs. To identify vaccine safety signals using VAERS data is critical and potentially impact millions of vaccine recipients, as VAERS is the front line for vaccine safety. In a sense, the proposed method for monitoring increase of AEs in VAERS is a signal detection method to provide early warning signals and generate meaningful hypotheses for further investigations.

Acknowledgement. The authors thank the referees, the associate editor and the editor for their constructive comments that substantially improved the presentation of this work. This work is supported in part by NIH Grants: R01AI130460, R01LM012607 and R01HD099348.

APPENDIX A: LIST OF ABBREVIATIONS

A list of abbreviations used in the main text was provided in Table 4.

APPENDIX B: ADDITIONAL SIMULATION STUDIES

In this appendix we present results from additional simulation studies. Tables 5 and 6 showed results from the same scenarios in Table 1, with rejection rates evaluated at the nominal levels of 0.01 and 0.0025, respectively. Table 7 showed results with $I = 100$ years, with rejection rates evaluated at the nominal levels of 0.05 and 0.001, respectively.

TABLE 4
Abbreviations used in the main text

AE	adverse event
BCPNN	Bayesian confidence propagation neutral network
CDC	Centers for Disease Control and Prevention
DTP	diphtheria, tetanus and acellular pertussis
FAERS	FDA Adverse Event Reporting System
FDA	Food and Drug Administration
FLU3	trivalent influenza virus vaccine
GBS	Guillain–Barré syndrome
MedDRA	Medical Dictionary for Regulatory Activities
MGPS	multiitem Gamma Poisson shrinker
MLRT	maximum likelihood ratio test
MMR	measles, mumps and rubella
MSAEFI	Monitoring System for Adverse Events Following Immunization
PT	preferred term
POR	reporting odds ratios
PRR	proportional reporting ratios
SOC	system organ class
VAERS	Vaccine Adverse Event Reporting System

TABLE 5

Empirical rejection rates (%) in 1000 *simulations of the proposed method to test the heterogeneity of reporting rates of AE over* $I = 25, 50$ *years, with the heterogeneity of the reporting rates* τ_j^2 *varying from* 0 *to* 0.1, *and the probability of observing zero event* w_j *increasing from* 0 *to* 0.8 *at the nominal level of* 0.01

	$I = 25$					$I = 50$				
	τ_j^2					τ_j^2				
w_j	0	0.025	0.05	0.075	0.1	0	0.025	0.05	0.075	0.1
0	0.6	14.2	44.2	68.8	83.4	0.5	31.0	70.7	94.2	98.2
0.2	0.5	11.5	35.2	59.3	74.2	1.1	22.8	63.2	86.8	94.5
0.4	0.9	9.1	26.2	44.2	57.1	0.6	17.6	48.8	76.0	89.9
0.6	0.1	4.9	17.7	29.6	41.5	0.6	12.7	37.1	55.7	72.5
0.8	0.2	4.0	7.5	12.8	18.7	0.7	4.9	13.8	30.8	41.6

TABLE 6

Empirical rejection rates (%) in 1000 *simulations of the proposed method to test the heterogeneity of reporting rates of AE over* $I = 25, 50$ *years, with the heterogeneity of the reporting rates* τ_j^2 *varying from* 0 *to* 0.1, *and the probability of observing zero event* w_j *increasing from* 0 *to* 0.8 *at the nominal level of* 0.0025

	$I = 25$					$I = 50$				
	τ_j^2					τ_j^2				
w_j	0	0.025	0.05	0.075	0.1	0	0.025	0.05	0.075	0.1
0	0.1	7.6	30.5	56.4	75.1	0.1	18.3	60.3	89.5	95.6
0.2	0.1	5.9	23.5	45.7	62.3	0.2	14.1	50.7	78.3	90.7
0.4	0.3	4.3	15.7	33.6	48.0	0.1	8.6	36.5	66.4	82.1
0.6	0.0	2.4	10.6	21.8	31.9	0.3	7.0	25.3	42.8	63.2
0.8	0.0	0.8	3.9	7.6	12.0	0.2	2.4	8.7	20.3	32.1

TABLE 7

Empirical rejection rates (%) in 1000 *simulations of the proposed method to test the heterogeneity of reporting rates of AE over* $I = 100$ *years, with the heterogeneity of the reporting rates* τ_j^2 *varying from* 0 *to* 0.1, *and the probability of observing zero event* w_j *increasing from* 0 *to* 0.8 *at the nominal levels of* 0.05 *and* 0.0025

	$\alpha = 0.05$					$\alpha = 0.0025$				
	τ_j^2					τ_j^2				
w_j	0	0.025	0.05	0.075	0.1	0	0.025	0.05	0.075	0.1
0	5.3	75.5	98.7	100.0	100.0	0.2	42.3	91.6	99.9	100.0
0.2	6.0	72.3	96.6	99.6	100.0	0.1	35.9	83.3	97.9	100.0
0.4	6.1	62.6	91.8	98.6	100.0	0.3	25.1	69.1	92.8	98.7
0.6	3.8	45.6	78.6	94.5	98.3	0.1	14.6	50.4	80.4	91.1
0.8	2.9	26.4	53.1	73.3	83.7	0.1	5.7	23.6	45.8	62.5

SUPPLEMENTARY MATERIAL

R functions and an example to implement the proposed method (DOI: 10.1214/20-AOAS1393SUPP; .zip). We provide R functions together with an example to implement the proposed composite likelihood ratio test.

REFERENCES

BATE, A., LINDQUIST, M., EDWARDS, I., OLSSON, S., ORRE, R., LANSNER, A. and DE FREITAS, R. M. (1998). A Bayesian neural network method for adverse drug reaction signal generation. *Eur. J. Clin. Pharmacol.* **54** 315–321.

CHEN, R. T., HABER, P. and MULLEN, J. R. (1995). Surveillance of the safety of simultaneous administration of vaccines: The CDC experience. *Ann. N.Y. Acad. Sci.* **754** 309–320.

CHEN, Y. and LIANG, K.-Y. (2010). On the asymptotic behaviour of the pseudolikelihood ratio test statistic with boundary problems. *Biometrika* **97** 603–620. MR2672486 https://doi.org/10.1093/biomet/asq031

CHEN, Y., NING, Y., HONG, C., and WANG, S. (2014). Semiparametric tests for identifying differentially methylated loci with case–control designs using illumina arrays. *Genet. Epidemiol.* **38** 42–50. https://doi.org/10.1002/gepi.21774

CHEN, Y., HONG, C., and RILEY, R. (2015). An alternative pseudolikelihood method for multivariate random-effects meta-analysis. *Stat. Med.* **34** 361–380. https://doi.org/10.1002/sim.6350

CHEN, Y., NING, J., NING, Y., LIANG, K.-Y., and BANDEEN-ROCHE, K. (2017). On pseudolikelihood inference for semiparametric models with boundary problems. *Biometrika* **104** 165–179. https://doi.org/10.1093/biomet/asw072

CHEN, R. T., RASTOGI, S. C., MULLEN, J. R., HAYES, S. W., COCHI, S. L., DONLON, J. A. and WASSILAK, S. G. (1994). The vaccine adverse event reporting system (VAERS). *Vaccine* **12** 542–550.

CHEN, R. T., GLASSER, J. W., RHODES, P. H., DAVIS, R. L., BARLOW, W. E., THOMPSON, R. S., MULLOOLY, J. P., BLACK, S. B., SHINEFIELD, H. R. et al. (1997). Vaccine safety datalink project: A new tool for improving vaccine safety monitoring in the United States. *Pediatrics* **99** 765–773.

CLAYTON, D. and KALDOR, J. (1987). Empirical Bayes estimates of age-standardized relative risks for use in disease mapping. *Biometrics* **43** 671–681.

COX, D. R. and REID, N. (2004). A note on pseudolikelihood constructed from marginal densities. *Biometrika* **91** 729–737. MR2090633 https://doi.org/10.1093/biomet/91.3.729

DU, J., CAI, Y., CHEN, Y. and TAO, C. (2016). Trivalent influenza vaccine adverse symptoms analysis based on MedDRA terminology using VAERS data in 2011. *J. Biomed. Semant.* **7** 13.

DUMOUCHEL, W. (1999). Bayesian data mining in large frequency tables, with an application to the FDA spontaneous reporting system. *Amer. Statist.* **53** 177–190.

EBERTH, J. M., KLINE, K. N., MOSKOWITZ, D. A., MONTEALEGRE, J. R. and SCHEURER, M. E. (2014). The role of media and the Internet on vaccine adverse event reporting: A case study of human papillomavirus vaccination. *J. Adolesc. Health* **54** 289–295.

ELLENBERG, S. S. and CHEN, R. T. (1997). The complicated task of monitoring vaccine safety. *Public Health Rep.* **112** 10.

EVANS, S., WALLER, P. C. and DAVIS, S. (2001). Use of proportional reporting ratios (PRRs) for signal generation from spontaneous adverse drug reaction reports. *Pharmacoepidemiol. Drug Saf.* **10** 483–486.

CENTERS FOR DISEASE CONTROL AND PREVENTION (1991). Pertussis vaccination: Acellular pertussis vaccine for the fourth and fifth doses of the DTP series; update to supplementary ACIP statement. *Morb. Mort. Wkly. Rep.* **40** 881–882.

CENTERS FOR DISEASE CONTROL AND PREVENTION (2016a). Seasonal influenza. Available at http://www.cdc.gov/flu/about/qa/disease.htm.

CENTERS FOR DISEASE CONTROL AND PREVENTION (2016b). Seasonal influenza vaccine safety: A summary for clinicians. Available at https://www.cdc.gov/flu/professionals/vaccination/vaccine_safety.htm.

CENTERS FOR DISEASE CONTROL AND PREVENTION (2017a). Estimated influenza illnesses, medical visits, hospitalizations and deaths averted by vaccination in the United States. Available at https://www.cdc.gov/flu/about/disease/2015-16.htm.

CENTERS FOR DISEASE CONTROL AND PREVENTION (2017b). Vaccine effectiveness. Available at https://www.cdc.gov/flu/about/qa/vaccineeffect.htm.

CENTERS FOR DISEASE CONTROL AND PREVENTION (2017c). Vaccine safety. Available at https://www.cdc.gov/vaccinesafety/ensuringsafety/monitoring/vaers/index.html.

CENTERS FOR DISEASE CONTROL AND PREVENTION (2017d). Vaccine safety datalink (VSD). Available at https://www.cdc.gov/vaccinesafety/ensuringsafety/monitoring/vsd/index.html.

GODAMBE, V. P. (1960). An optimum property of regular maximum likelihood estimation. *Ann. Math. Stat.* **31** 1208–1211. MR0123385 https://doi.org/10.1214/aoms/1177705693

GOLDMAN, G. (2013). Comparison of VAERS fetal-loss reports during three consecutive influenza seasons: Was there a synergistic fetal toxicity associated with the two-vaccine 2009/2010 season? *Hum. Exp. Toxicol.* **32** 464–475.

GREENWOOD, P. E. and NIKULIN, M. S. (1996). *A Guide to Chi-Squared Testing*. *Wiley Series in Probability and Statistics*: *Applied Probability and Statistics*. Wiley, New York. MR1379800

GROHSKOPF, L. A., SOKOLOW, L. Z., OLSEN, S. J., BRESEE, J. S., BRODER, K. R. and KARRON, R. A. (2015). Prevention and control of influenza with vaccines: Recommendations of the advisory committee on immunization practices, United States, 2015–16 influenza season. *Morb. Mort. Wkly. Rep.* **64** 818–825.

HABER, P., DESTEFANO, F., ANGULO, F. J., ISKANDER, J., SHADOMY, S. V., WEINTRAUB, E. and CHEN, R. T. (2004). Guillain–Barré syndrome following influenza vaccination. *JAMA* **292** 2478–2481.

HUANG, L., ZALKIKAR, J. and TIWARI, R. C. (2011). A likelihood ratio test based method for signal detection with application to FDA's drug safety data. *J. Amer. Statist. Assoc.* **106** 1230–1241. MR2896832 https://doi.org/10.1198/jasa.2011.ap10243

HUANG, L., ZHENG, D., ZALKIKAR, J. and TIWARI, R. (2017). Zero-inflated Poisson model based likelihood ratio test for drug safety signal detection. *Stat. Methods Med. Res.* **26** 471–488. MR3592736 https://doi.org/10.1177/0962280214549590

HUANG, J., NING, Y., CAI, Y., LIANG, K.-Y. and CHEN, Y. (2020). Composite likelihood inference under boundary conditions. *Statist. Sinica* **30** 1005–1025.

HUANG, J., CAI, Y., DU, J., LI, R., ELLENBERG, S. S., HENNESSY, S., TAO, C. and CHEN, Y. (2021). Supplement to "Monitoring vaccine safety by studying temporal variation of adverse events using vaccine adverse event reporting system." https://doi.org/10.1214/20-AOAS1393SUPP

IQBAL, S., LI, R., GARGIULLO, P. and VELLOZZI, C. (2015). Relationship between Guillain–Barré syndrome, influenza-related hospitalizations, and influenza vaccine coverage. *Vaccine* **33** 2045–2049.

LASKY, T., TERRACCIANO, G. J., MAGDER, L., KOSKI, C. L., BALLESTEROS, M., NASH, D., CLARK, S., HABER, P., STOLLEY, P. D. et al. (1998). The Guillain–Barré syndrome and the 1992–1993 and 1993–1994 influenza vaccines. *N. Engl. J. Med.* **339** 1797–1802.

LIEVANO, F., GALEA, S. A., THORNTON, M., WIEDMANN, R. T., MANOFF, S. B., TRAN, T. N., AMIN, M. A., SEMINACK, M. M., VAGIE, K. A. et al. (2012). Measles, mumps, and rubella virus vaccine (M-M-R^{TM} II): A review of 32 years of clinical and postmarketing experience. *Vaccine* **30** 6918–6926.

LINDSAY, B. G. (1988). Composite likelihood methods. In *Statistical Inference from Stochastic Processes* (*Ithaca*, *NY*, 1987). *Contemp. Math.* **80** 221–239. Amer. Math. Soc., Providence, RI. MR0999014 https://doi.org/10.1090/conm/080/999014

MCNEIL, M. M., GEE, J., WEINTRAUB, E. S., BELONGIA, E. A., LEE, G. M., GLANZ, J. M., NORDIN, J. D., KLEIN, N. P., BAXTER, R. et al. (2014). The vaccine safety datalink: Successes and challenges monitoring vaccine safety. *Vaccine* **32** 5390–5398.

MEDDRA, M. (2017). Introductory Guide MedDRA Version 20.1. MedDRA Maintenance and Support Services Organization.

PEARSON, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Philos. Mag.* **50** 157–175.

ROTHMAN, K. J., LANES, S. and SACKS, S. T. (2004). The reporting odds ratio and its advantages over the proportional reporting ratio. *Pharmacoepidemiol. Drug Saf.* **13** 519–523.

SHIMABUKURO, T. T., NGUYEN, M., MARTIN, D. and DESTEFANO, F. (2015). Safety monitoring in the vaccine adverse event reporting system (VAERS). *Vaccine* **33** 4398–4405.

SINGLETON, J. A., LLOYD, J. C., MOOTREY, G. T., SALIVE, M. E., CHEN, R. T., ELLENBERG, S., RASTOGI, S., KRUEGER, C., BRAUN, M. et al. (1999). An overview of the vaccine adverse event reporting system (VAERS) as a surveillance system. *Vaccine* **17** 2908–2917.

THOMPSON, K. M., GELLIN, B. G., HINMAN, A. R. and ORENSTEIN, W. A. (2018). The National Vaccine Advisory Committee at 30: Impact and opportunity. *Vaccine* **36** 1330–1344. https://doi.org/10.1016/j.vaccine.2018.01.068

VARIN, C., REID, N. and FIRTH, D. (2011). An overview of composite likelihood methods. *Statist. Sinica* **21** 5–42. MR2796852

VELLOZZI, C., BURWEN, D. R., DOBARDZIC, A., BALL, R., WALTON, K. and HABER, P. (2009). Safety of trivalent inactivated influenza vaccines in adults: Background for pandemic influenza vaccine safety monitoring. *Vaccine* **27** 2114–2120.

WEBER, J. C. P. (1984). Epidemiology of adverse reactions to nonsteroidal anti-inflammatory drugs. In *Advances in Inflammation Research* 1–7.

ZHAO, Y., YI, M. and TIWARI, R. C. (2018). Extended likelihood ratio test-based methods for signal detection in a drug class with application to FDA's adverse event reporting system database. *Stat. Methods Med. Res.* **27** 876–890. MR3767629 https://doi.org/10.1177/0962280216646678

ZHOU, W., POOL, V., ISKANDER, J. K., ENGLISH-BULLARD, R., BALL, R., WISE, R. P., HABER, P., PLESS, R. P., MOOTREY, G. et al. (2003). Surveillance for safety after immunization: Vaccine adverse event reporting system (VAERS)—United States, 1991–2001. *MMWR Surveill. Summ.* **52** 1–24.